

تشخیص انجمن‌ها در شبکه‌های اجتماعی با الگوریتم تخمین توزیع

محسن قاسمی پارسا^۱ و ناصر مزینی^۲

دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران: m_parsa@comp.iust.ac.ir

دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران: mozayani@iust.ac.ir

چکیده:

شبکه‌های اجتماعی دارای واحدهای بنیادینی به نام انجمن می‌باشند که فهم ساختار آن‌ها نقش مهمی در فهم ساختار یک شبکه‌ی بزرگ ایفا می‌کند. انجمن‌ها گروه‌هایی از رأس‌ها می‌باشند که اتصال‌های زیادی در درون و اتصال‌های کمی با بیرون خود دارند. تشخیص انجمن‌ها در یک شبکه، می‌تواند به عنوان یک مسئله‌ی بهینه‌سازی مدل شود. با توجه به اینکه برای اندازه‌گیری کیفیت هر یک از افرازهای رأس‌های شبکه، می‌توان از توابعی نظیر پیمانه‌ای بودن استفاده نمود، از این توابع به عنوان تابع هدف در اینگونه مسائل بهینه‌سازی استفاده می‌شود. در این نوشتار از الگوریتم تخمین توزیع با فرض استقلال بین ژن‌ها برای تشخیص انجمن‌ها استفاده شده است. الگوریتم تخمین توزیع یکی از الگوریتم‌های تکاملی است که در آن به جای استفاده از عملگرهای ژنتیکی نظیر بازترکیب و جهش، ابتدا توزیع احتمال نمونه‌های انتخاب شده از نسل فعلی تخمین زده می‌شود و بر اساس این توزیع، نسل بعدی ایجاد می‌گردد. در این نوشتار ضمن ارزیابی این روش تشخیص انجمن‌ها با مجموعه داده‌های ساختگی و مجموعه داده‌های واقعی، نتایج بدست آمده با دیگر روش‌های تشخیص انجمن‌ها مقایسه شده‌است.

واژه‌های کلیدی: شبکه‌های اجتماعی، تشخیص انجمن‌ها، محاسبات تکاملی، الگوریتم تخمین توزیع

۱- مقدمه

ظهور شبکه‌های اجتماعی آنلاین یکی از مهم‌ترین رخداد‌های دهه‌ی اخیر بوده است. محبوبیت این شبکه‌ها نظیر فیسبوک^۳، توئیتر^۴، لینکدین^۵ بطور روزافزون در حال افزایش است. این شبکه‌ها منبع عظیمی از داده‌هاست که شامل محتوا^۶ و پیوند^۷ می‌شوند و می‌توانند مورد تجزیه و تحلیل قرار گیرند. داده‌های پیوند، بصورت گراف بازنمایی می‌شود که در آن رأس‌های گراف، موجودیت‌ها و یال‌ها معرف ارتباط بین موجودیت‌ها می‌باشد. در حالی که محتوا شامل متن، تصاویر و دیگر داده‌های چندرسانه‌ای می‌شود. بطور کلی یک شبکه‌ی اجتماعی به عنوان شبکه‌ای از تعاملات^۸ یا رابطه‌ها^۱ شناخته می‌شود.

۱- دانشجوی کارشناسی ارشد دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

۲- دانشیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

³ Facebook

⁴ Twitter

⁵ LinkedIn

⁶ Content

⁷ Linkage

⁸ Interactions

که رأس‌ها شامل کنش‌گرها^۲ و یال‌ها شامل رابطه‌ها و تعاملات بین این کنش‌گرها می‌شود. یک تعمیم از ایده شبکه‌های اجتماعی، شبکه‌های اطلاعاتی^۳ می‌باشد که در آن رأس‌ها می‌تواند شامل کنش‌گرها یا هر موجودیت دیگر باشد و یال‌ها به معنی رابطه‌های بین آن‌هاست [۱].

یکی از مهم‌ترین مسائل مطرح در حوزه شبکه‌های اجتماعی، تشخیص انجمن‌ها^۴ می‌باشد. مسئله تشخیص انجمن‌ها مشخصاً مرتبط با خوشه‌بندی^۵ می‌باشد و بخش‌هایی از شبکه که به لحاظ پیوند بین موجودیت‌ها چگالی بالایی دارند را مشخص می‌کند. این مبحث به مسئله عمومی افراز گراف^۶ مرتبط می‌شود، که با توجه به پیوندهای موجود بین رأس‌ها، شبکه را به قسمت‌هایی با چگالی بالا افراز می‌کند [۲]. با توجه به اینکه انجمن‌ها واحدهای بنیادینی در شبکه‌های اجتماعی می‌باشند، فهم ساختار انجمن‌ها نقش مهمی در فهم ساختار یک شبکه بزرگ ایفا می‌کند.

روش‌های زیادی برای تشخیص انجمن‌ها در شبکه‌های پیچیده^۷ وجود دارد که در ادامه به برخی از این روش‌ها اشاره می‌شود. الگوریتم Girvan-Newman (GN) [۴-۵] یک روش سلسله‌مراتبی است که یال‌های شبکه را با توجه به مقدار مابین بودن^۸ حذف می‌نماید. این مقدار، نمایان‌گر تعداد دفعاتی است که یک یال در مسیر کوتاه‌ترین مسیر بین هر جفت از رأس‌ها قرار می‌گیرد. فرآیند حذف یال‌ها تا زمانی که معیاری مانند پیمانهای بودن افزایش پیدا نکند، ادامه می‌یابد. الگوریتم Isomap [۶] برای تشخیص انجمن‌های یک شبکه، سعی در فشرده‌سازی بهینه‌ی اطلاعات ساختار شبکه را دارد. الگوریتم Clauset-Newman-Moor (CNM) [۷] ابتدا با رأس‌های بدون یال شبکه آغاز می‌کند و سپس شروع به اضافه‌نمودن یال‌های شبکه اصلی با توجه به افزایش معیار پیمانهای بودن می‌کند. الگوریتم Louvain [۸] سعی در بهینه‌سازی محلی تابع پیمانهای بودن دارد. الگوریتم GA-Net [۹] با استفاده از الگوریتم ژنتیک و تابع هدف امتیاز انجمن^۹ سعی در تشخیص انجمن‌ها دارد. Li و Song [۱۰] نیز از محاسبات تکاملی برای تشخیص انجمن‌ها استفاده نموده‌اند، با این تفاوت که الگوریتم تخمین توزیع ECGA^{۱۰} که وابستگی بین ژن‌ها نیز در آن لحاظ شده را بکار گرفته‌اند.

الگوریتم‌های تخمین توزیع^{۱۱} به آن دسته از تکنیک‌های بهینه‌سازی اتفاقی گفته می‌شود که فضای جواب‌ها را بوسیله‌ی ایجاد و نمونه‌گیری از مدل‌های احتمالی برای رسیدن به جواب‌های دلخواه، کاوش می‌کنند [۱۱]. به بیان دیگر الگوریتم‌های تخمین توزیع، آن دسته از الگوریتم‌های تکاملی می‌باشند که به جای استفاده از عملگرهای ژنتیکی نظیر باز ترکیب و جهش در هریک از نسل‌ها، مدل احتمالی برخی نمونه‌های انتخاب شده از نسل فعلی را محاسبه کرده و با نمونه‌گیری از این مدل احتمالی، نسل بعد را ایجاد می‌کنند. الگوریتم‌های تخمین توزیع، بسته به اینکه متغیرها (در این جا ژن‌ها) مستقل از هم باشند یا وابسته، بازنمایی جواب‌ها بصورت دودویی، عدد صحیح یا عدد حقیقی باشد، از چه مدل احتمالی (نظیر مارکف، شبکه بیزین، خوشه‌های مستقل و درختی) استفاده کنند، به الگوریتم‌های متنوعی تقسیم‌بندی می‌شوند که در [۱۱] بطور مفصل بیان شده است. در این نوشتار برای تشخیص انجمن‌ها در شبکه‌های اجتماعی از الگوریتم تخمین توزیع با فرض

¹ Relationships

² Actors

³ Information networks

⁴ Community detection

⁵ Clustering

⁶ Graph-partitioning

⁷ Complex networks

⁸ Betweenness

⁹ Community score

¹⁰ Extended Compact Genetic Algorithm

¹¹ Estimation of distribution algorithms (EDAs)

مستقل بودن ژن‌ها از یکدیگر استفاده شده است. فرض استقلال بین ژن‌ها باعث سادگی روش و حذف سربار محاسباتی ناشی از الگوریتم‌های یادگیری پیوند^۱ می‌شود.

در ادامه و در بخش ۲، مسئله تشخیص انجمن‌ها بوسیله بهینه‌سازی تعریف می‌شود و دو تابع هدف که می‌تواند برای بهینه‌سازی افراز انجمن‌ها استفاده گردد، شرح داده می‌شود. بخش ۳ به تشریح الگوریتم می‌پردازد و در بخش ۴ آزمایش‌ها و نتایج بدست آمده ارائه می‌گردد، ضمن اینکه در بخش ۵ نتیجه‌گیری کلی و کارهای آینده بیان می‌شود.

۲- تعریف مسئله

ساختار شبکه‌های اجتماعی را می‌توان با گراف بازنمایی نمود، که رأس‌های گراف معرف موجودیت‌ها و یال‌های آن معرف روابط بین آن‌ها می‌باشد. تشخیص انجمن‌ها در شبکه‌های اجتماعی به معنی خوشه‌بندی رأس‌ها در گروه‌هایی است که رأس‌های داخل گروه، یال‌های زیادی بین خود داشته باشند و بین گروه‌ها یال‌های کمی وجود داشته باشد. تشخیص انجمن‌ها در شبکه‌های اجتماعی را می‌توان به عنوان یک مسئله بهینه‌سازی با توجه به یک تابع هدف^۲ مدل کرد، که این مسئله در دسته الگوریتم‌های NP-کامل قرار می‌گیرد [۱۲]. برای اینگونه مسائل بهینه‌سازی، همیشه الگوریتم‌های تکاملی گزینه‌ی مطرحی می‌باشند. در ادامه این بخش یک نمونه از معیارهایی که به عنوان تابع هدف برای افراز گراف استفاده می‌گردد، ارائه می‌شود.

۲-۱- تابع پیمان‌های بودن

این معیار ابتدا در [۵] مطرح گردید که برای توقف فرآیند حذف اتصالات شبکه استفاده می‌شود. پیمان‌های بودن (Q) به عنوان معیار مناسبی برای تشخیص اینکه یک افراز با چه کیفیتی رأس‌ها را خوشه‌بندی می‌کند، شناخته می‌شود. فرض کنید که A_{vw} یک عنصر از ماتریس مجاورت شبکه باشد به نحوی که اگر رأس‌های v و w به هم متصل باشند، مقدار یک و در غیر این صورت مقدار صفر را دارا باشد. m معرف تعداد یال‌های شبکه می‌باشد $(m = \frac{1}{2} \sum_{vw} A_{vw})$ ، ضمن اینکه رأس‌ها به انجمن‌هایی تقسیم می‌شوند و c_v معرف انجمن رأس v و k_v درجه رأس v می‌باشد و بدین صورت تعریف می‌گردد:

$$k_v = \sum_w A_{vw} \quad (1)$$

تابع پیمان‌های بودن بصورت زیر تعریف می‌گردد:

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w) \quad (2)$$

که در آن خروجی تابع δ در صورتی که دو ورودی‌اش برابر با هم باشند مقدار ۱ و در غیر این صورت صفر است. می‌توان Q را بدین صورت نیز بازنویسی کرد:

$$Q = \sum_{i=1}^{n_c} (e_{ii} - a_i^2) \quad (3)$$

¹ Linkage learning

² Objective function

که در آن n_c تعداد انجمن‌ها می‌باشد و e_{ii} و a_i بدین صورت بدست می‌آیند:

$$e_{ii} = \frac{1}{2m} \sum_{vw} A_{vw} \delta(c_v, c_w) \quad (4)$$

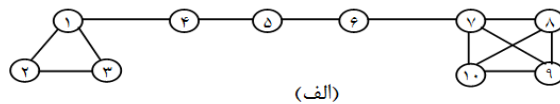
$$a_i = \frac{1}{2m} \sum_v k_v \delta(c_v, i) \quad (5)$$

۳- روش پیشنهادی

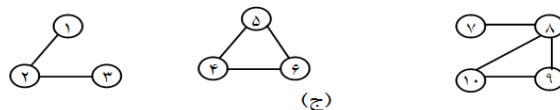
در این بخش روش پیشنهادی ما برای تشخیص انجمن‌ها تشریح می‌گردد که شامل نحوه‌ی بازنمایی و تشریح الگوریتم ارائه شده می‌باشد.

۳-۱- بازنمایی جواب

بازنمایی جواب‌ها در اینجا مانند بازنمایی ژنتیکی استفاده شده در [۹] می‌باشد که به بازنمایی مجاورت مبتنی بر موقعیت^۱ معروف است. در این روش یک عضو جمعیت شامل N ژن (g_1 تا g_N) می‌شود که هر یک از ژن‌ها می‌تواند مقدار $j \in \{1, \dots, N\}$ را اختیار کند. ژن‌ها و مقادیرشان رأس‌های گراف $G = (V, E)$ را بازنمایی می‌کنند. هر یک از اعضای جمعیت، یک گراف را نشان می‌دهد، بدین صورت که اگر ژن i مقدار j را اختیار کند، بین رأس‌های i و j یک یال قرار می‌گیرد و بدین معنی است که رأس i و j در یک خوشه قرار دارند. در این بازنمایی مرحله دیکد کردن با بدست آوردن اجزای همبند گراف ایجاد شده، در زمان خطی انجام می‌شود. مزیت اصلی این روش بازنمایی، تعیین تعداد خوشه‌ها بصورت خودکار می‌باشد و نیازی به مشخص بودن تعداد انجمن‌ها از قبل نمی‌باشد. یک نمونه از این بازنمایی در شکل ۱ بیان شده است. در قسمت (الف) شکل ۱، یک مثال از گراف شبکه نشان داده شده است. در قسمت (ب) یک جواب برای افراز این شبکه آمده است و در قسمت (ج) گراف ناشی از این جواب نشان داده شده است. برای دیکد کردن این جواب کافی است اجزای همبند گراف (ج) را با زمان خطی بدست آورد.



| موقعیت | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ |
|--------|---|---|---|---|---|---|---|---|----|----|
| مقدار | ۲ | ۱ | ۲ | ۵ | ۶ | ۴ | ۸ | ۹ | ۱۰ | ۸ |



شکل ۱: بازنمایی افراز رأس‌ها. (الف) گراف شبکه (ب) بازنمایی جواب (ج) گراف ناشی از بازنمایی

¹ Locus-based adjacency representation

۳-۲- جمعیت اولیه

برای ایجاد جمعیت اولیه در فرآیند تکامل، هریک از ژن‌ها می‌تواند شماره‌ی رأس یکی از همسایگان خود و یا شماره‌ی خود را اختیار کند. این نوع ایجاد جمعیت اولیه، بدلیل هرس فضای جستجو بکار می‌رود.

۳-۳- الگوریتم ارائه شده

برای تشخیص انجمن‌ها در شبکه‌های اجتماعی از الگوریتم تخمین توزیع با این فرض که ژن‌ها از یکدیگر مستقل هستند استفاده شده است. در هریک از نسل‌ها، تعدادی از جمعیت فعلی انتخاب می‌گردد و توزیع احتمال هریک از ژن‌ها محاسبه می‌گردد. این توزیع احتمال بر اساس مقادیر مختلفی که آن ژن در نمونه‌های انتخاب شده دارد، محاسبه می‌گردد. این توزیع‌ها به عنوان مدل احتمالاتی نسل فعلی در نظر گرفته می‌شود و با نمونه‌گیری از این توزیع‌های احتمال، نسل بعدی ایجاد می‌گردد. الگوریتم ۱ بصورت رسمی روش ارائه شده را بیان می‌کند.

الگوریتم ۱: تشخیص انجمن‌ها با الگوریتم تخمین توزیع

۱- جمعیت اولیه را بدین صورت ایجاد کن: مقدار ژن A می‌تواند خود A یا یکی از همسایه‌های رأس A در شبکه باشد

۲- مقدار برازش اعضای جمعیت را محاسبه کن

۳- تعداد n عضو از کل N عضو جمعیت را به روش تورنومنت انتخاب کن

۴- توزیع احتمال هریک از ژن‌ها را با توجه به مقادیر مختلفی که در n عضو انتخابی دارند، محاسبه کن

۵- باتوجه به توزیع‌های احتمال بدست آمده، برای ایجاد نسل بعد از این توزیع‌ها نمونه‌گیری کن

۶- با احتمال پایین P_m نمونه‌های بدست آمده در مرحله ۵ را تغییر بده (در این عمل شبه جهش هریک از ژن‌ها مانند مرحله ۱ فقط می‌تواند مقداری برابر شماره رأس متناظر با آن ژن یا همسایه‌های آن ژن را اختیار کند)

۷- خروجی مرحله ۶ و عضو از نسل قبل با بیشترین برازش را به عنوان اعضای جمعیت فعلی در نظر بگیر

۸- در صورتی که بهترین عضو نسل قبل در g نسل قبلی تغییری نکرده باشد، الگوریتم را خاتمه بده. در غیر اینصورت به مرحله ۲ برو

در مرحله ۶ و برای جلوگیری از همگرایی زودرس^۱ یک عمل شبه جهش انجام شده است. با افزایش مقدار P_m ، توانایی الگوریتم در اکتشاف^۲ فضای جستجو افزایش می‌یابد.

باتوجه به اینکه در این روش از عملگر ژنتیکی باز ترکیب^۳ استفاده نمی‌شود و اعضا صرفاً با نمونه‌گیری از مدل احتمالاتی ایجاد می‌گردند، مزیت این الگوریتم نسبت به دیگر الگوریتم‌های تکاملی نظیر [۹] و [۱۳] سرعت ایجاد نسل جدید در آن است. مزیت دیگر این الگوریتم، سرعت همگرایی آن است که به خاطر استفاده از الگوریتم تخمین توزیع میسر شده است. مزیت این روش نسبت به دیگر روش‌هایی نظیر [۱۰] که وابستگی بین ژن‌ها را در ایجاد مدل احتمالاتی لحاظ می‌کنند، سادگی و سریع‌تر شدن فرآیند ایجاد مدل می‌باشد.

¹ Premature convergence

² Exploration

³ Recombination

۴- آزمایش‌ها

در این بخش، روش ارائه شده ارزیابی می‌شود و با دیگر روش‌ها مقایسه می‌گردد. ما از کتابخانه ارائه شده با نام SNAP^۱ برای اجرای سه روش Girvan-Newman [۴] (بطور اختصار GN)، Clauset-Newman-Moore [۷] (بطور اختصار CNM) و Infomap [۶] استفاده نموده‌ایم. همه‌ی آزمایش‌ها در کامپیوتری با پردازنده‌ی Core2 Duo 1.83GHz و مقدار حافظه اصلی ۲ گیگابایت انجام گردیده است. از دو دسته مجموعه داده‌ی واقعی و ساختگی^۲ که برچسب‌های اصلی هریک از رأس‌ها را نیز با خود دارند، استفاده نموده‌ایم که در ادامه این دو دسته تشریح می‌گردد. ضمن اینکه برای ارزیابی چگونگی تشخیص انجمن از معیار اطلاعات متقابل^۳ [۱۴] استفاده شده است.

در این آزمایش‌ها از تابع پیمانه‌ای بودن استفاده گردیده است. پارامترهای الگوریتم بدین صورت تنظیم شده‌اند، اندازه جمعیت (N) برابر ۳۰۰، احتمال جهش (P_m) برابر ۰,۰۰۲، روش انتخاب تورنومنت با $q=10$ و انتخاب ۵۰ عضو برای بدست آوردن مدل احتمالاتی (n). روی مجموعه داده‌های واقعی به تعداد ۱۰ بار روش‌مان را آزمایش نمودیم و روی هریک از ۳۰ سری مجموعه داده‌ی ساختگی نیز هرکدام یک بار روش‌مان را آزمایش نموده‌ایم.

۴-۱- مجموعه داده‌های ساختگی

ما از مجموعه داده پیشنهاد شده توسط Girvan و Newman در [۴] برای مقایسه با روش خودمان استفاده نموده‌ایم. این مجموعه داده شامل ۱۲۸ رأس می‌شود که رأس‌ها بطور مساوی در ۴ انجمن قرار می‌گیرند. برای ایجاد این مجموعه داده بایستی ضریب آمیختن^۴ را تعیین کنیم. این پارامتر نسبت بین تعداد اتصالاتی که یک رأس به داخل انجمنش دارد به کل اتصالاتی که یک رأس دارد را تعیین می‌کند. هرچه مقدار این پارامتر از صفر به یک نزدیک‌تر شود، تعداد اتصالات یک رأس به رأس‌های دیگر انجمن‌ها بیشتر شده و تشخیص انجمن‌ها در آن شبکه مشکل‌تر می‌شود. ما ۸ سری از این مجموعه داده ایجاد نموده‌ایم که ضریب آمیختن هر سری به ترتیب ۰,۱-۰,۲-۰,۳-۰,۴-۰,۵-۰,۶-۰,۷-۰,۸ می‌باشد و هر سری شامل ۳۰ مجموعه داده می‌شود.

در شکل ۲ نتیجه‌ی اعمال روش‌های مختلف تشخیص انجمن‌ها روی مجموعه داده‌ی ساختگی ذکر شده با ضرایب آمیختن مختلف نشان داده شده است. Ours روش ارائه شده‌ی ما می‌باشد. همانطور که در شکل ۲ مشاهده می‌کنید، نتایج ما وقتی که ضریب آمیختن کمتر از ۰,۶ است، بهتر از باقی روش‌هاست.

۴-۲- مجموعه داده‌های واقعی

ما از سه مجموعه داده‌ی فوتبال آمریکایی^۵، کتاب‌های سیاسی^۶ و دلفین‌ها^۷ استفاده نموده‌ایم. مجموعه داده‌ی فوتبال آمریکایی [۴] شامل شبکه‌ای از بازی‌های فوتبال آمریکایی بین یک‌سری کالج‌ها در سال ۲۰۰۰ می‌شود. این شبکه شامل ۱۱۵ رأس می‌شود که به ۱۲ گروه تقسیم می‌شوند که هریک از گروه‌ها نشان می‌دهد که یک تیم متعلق به کدام کنفرانس می‌شود. مجموعه داده‌ی کتاب‌های سیاسی شامل ۱۰۵ رأس می‌شود که رأس‌ها کتاب‌هایی درباره سیاست‌های آمریکا را

^۱ <http://snap.stanford.edu>

^۲ Synthetic

^۳ Normalized Mutual Information(NMI)

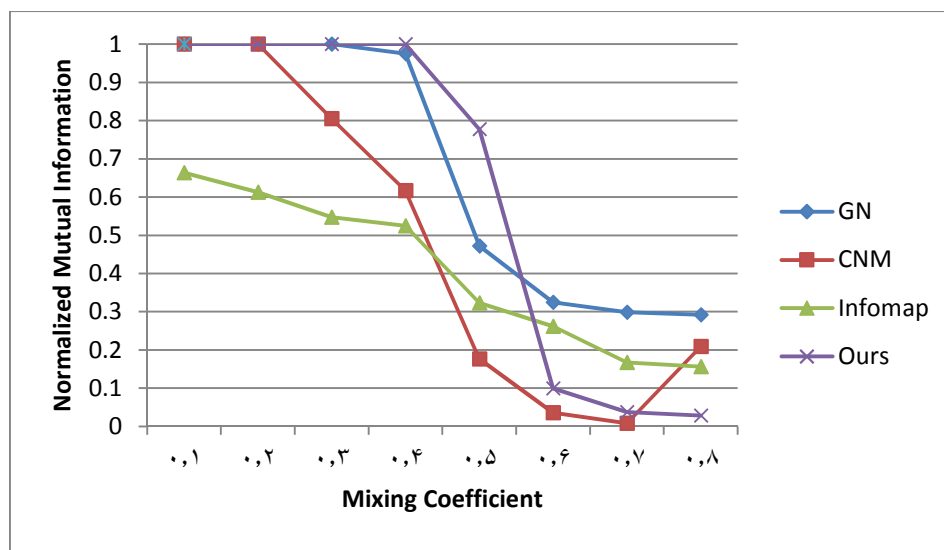
^۴ Mixing coefficient

^۵ American football games between Division IA colleges

^۶ Books about US politics

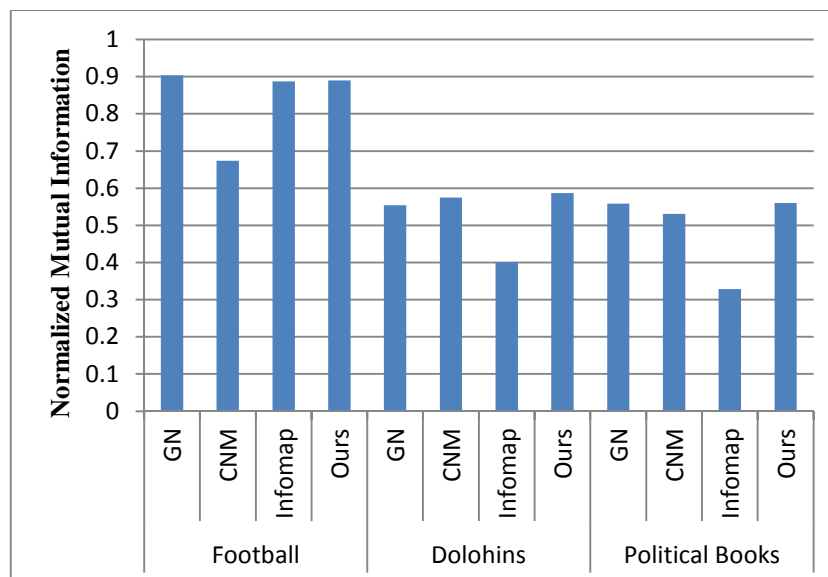
^۷ Associations between 62 dolphins

بازنمایی می‌کنند که بصورت آنلاین از آمازون خریداری شده‌اند. یال‌ها، خریده‌های همزمان کتاب‌ها توسط خریداران یکسان را بازنمایی می‌کند. رأس‌ها (کتاب‌ها) به ۳ گروه آزاد، خنثی و محافظه‌کار^۱ تقسیم می‌گردد. این تقسیم بندی توسط Newman [۱۶] انجام شده است. مجموعه داده‌ی دلفین‌ها [۱۷] شامل شبکه‌ای از ارتباط بین ۶۲ دلفین در منطقه‌ای واقع در نیوزلند می‌شود که به دو گروه تقسیم می‌شوند.



شکل ۲- نتایج حاصل از اعمال روش‌های مختلف روی مجموعه داده‌ی ساختگی Girvan-Newman و Ours روش ارائه شده می‌باشد

شکل ۳ نتایج حاصل از اعمال روش‌های مختلف تشخیص انجمن‌ها روی سه مجموعه داده‌ی واقعی را نشان می‌دهد. Ours روش ارائه شده‌ی ما می‌باشد. نتایج حاصل از روش ما در دو مجموعه داده‌ی دلفین‌ها و کتاب‌های سیاسی بهتر از دیگر روش‌ها و در مجموعه داده‌ی فوتبال آمریکایی به بهترین روش بسیار نزدیک می‌باشد.



شکل ۳- نتایج حاصل از اعمال روش‌های مختلف روی مجموعه داده‌های واقعی. Ours روش ارائه شده می‌باشد

¹ Liberal, Neutral, or Conservative

۵- نتیجه‌گیری و کارهای آینده

تشخیص انجمن‌ها یک مسئله‌ی مطرح در حوزه شبکه‌های اجتماعی می‌باشد. انجمن‌ها واحدهای بنیادینی می‌باشند که تشخیص آن‌ها به ما در فهم ساختار شبکه کمک شایانی می‌نماید. روش‌های مختلف و متنوعی برای تشخیص انجمن وجود دارد. در این نوشتار ما از الگوریتم تخمین توزیع با تابع هدف پیمان‌ای بودن برای تشخیص انجمن‌ها بهره بردیم و روش‌مان را روی مجموعه داده‌های مختلف آزمایش نمودیم و نتایج قابل قبولی را گرفتیم. مزیت روش ما نسبت به روش‌های حرصانه، جستجوی گسترده‌تر در فضای همه افرازهای شبکه می‌باشد و مزیت الگوریتم ما نسبت به روش‌های مبتنی بر الگوریتم ژنتیک، سرعت همگرایی می‌باشد. با توجه به اینکه در الگوریتم ما ژن‌ها مستقل از هم فرض شده‌اند، سر بار یادگیری پیوند در روش ما وجود ندارد. آزمایش‌ها نشان از توانمندی الگوریتم ما در تشخیص انجمن‌ها دارد.

برای کارهای آینده ایده‌هایی داریم که در ادامه می‌آیند. پرداختن به شبکه‌هایی که انجمن‌های هم‌پوشان دارند، استفاده از دیگر روش‌های تک متغیره‌ی الگوریتم تخمین توزیع، آزمایش روی مجموعه داده‌ی ساختگی LFR، استفاده از تابع امتیاز انجمن به عنوان تابع هدف و ایده گرفتن از روش‌های غیر تکاملی برای بهتر نمودن جواب بدست آمده از فرآیند تکامل، نمونه‌هایی از این ایده‌ها می‌باشد.

مراجع

- [1] C. Aggarwal, "An Introduction to Social Network Data Analytics," in *Social Network Data Analytics*, C. C. Aggarwal, Ed., ed: Springer US, 2011, pp. 1-15.
- [2] S. Parthasarathy, et al., "Community Discovery in Social Networks: Applications, Methods and Emerging Trends," in *Social Network Data Analytics*, C. C. Aggarwal, Ed., ed: Springer US, 2011, pp. 79-113.
- [3] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, pp. 75-174, 2010.
- [4] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 7821-7826, 2002.
- [5] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, p. 026113, 2004.
- [6] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 1118-1123, 2008.
- [7] A. Clauset, et al., "Finding community structure in very large networks," *Physical review E*, vol. 70, p. 066111, 2004.
- [8] V. D. Blondel, et al., "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, 2008.
- [9] C. Pizzuti, "GA-Net: A Genetic Algorithm for Community Detection in Social Networks," in *Parallel Problem Solving from Nature – PPSN X*, vol. 5199, G. Rudolph, et al., Eds., ed: Springer Berlin Heidelberg, 2008, pp. 1081-1090.
- [10] J. Li and Y. Song, "Community detection in complex networks using extended compact genetic algorithm," *Soft Computing*, vol. 17, pp. 925-937, 2013/06/01 2013.
- [11] M. Hauschild and M. Pelikan, "An introduction and survey of estimation of distribution algorithms," *Swarm and Evolutionary Computation*, vol. 1, pp. 111-128, 2011.
- [12] B. Karrer, et al., "Robustness of community structure in networks," *Physical review E*, vol. 77, p. 046119, 2008.
- [13] M. Tasgin, et al., "Community detection in complex networks using genetic algorithms," *arXiv preprint arXiv:0711.0491*, 2007.
- [14] L. N. F. Ana and A. K. Jain, "Robust data clustering," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, pp. II-128-II-133 vol.2.
- [15] C. Pizzuti, "Mesoscopic analysis of networks with genetic algorithms," *World Wide Web*, vol. 16, pp. 545-565, 2013/11/01 2013.
- [16] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 8577-8582, 2006.
- [17] D. Lusseau, et al., "The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, pp. 396-405, 2003.